Automated Highlights Generator for DOTA 2 Game Using Audio-Visual Framework

A dissertation submitted for the Degree of Master of Computer Science

Sanjeewa K.W.C.K.

University of Colombo School of Computing

2021

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information that have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

Student Name: Sanjeewa K.W.C.K.

Registration Number: 2018/MCS/081

Index Number: 18440814

2021/11/29

Signature of the Student & Date

This is to certify that this thesis is based on the work of Mr.Sanjeewa K.W.C.K. under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by,

Supervisor Name: MGNAS Fernando 29/11/2021

Signature of the Supervisor & Date

ACKNOWLEDGEMENTS

It is a pleasure to thank those who made this research work possible. I am grateful to my supervisor Dr. Noel Fernando of the University of Colombo, School of Computing for his valuable support to make my dissertation and the research project a success. The door to Dr. Fernando's office was always open and available for remote meetings during the pandemic whenever I was in a trouble or had a question about my research or writing. He consistently allowed this study to be my work but directed me in the right direction whenever he thought I needed it. His great guidance, remarkable patience, timely advice, well-organized meetings, excellent dissertation ideas, and confidence pointed me to come up with a successful research project. There are not enough kind words to express my appreciation for his involvement in this research work.

I want to express my gratitude to all the staff members and other members of the University of Colombo, School of Computing for the support given throughout the postgraduate journey.

Finally, I must direct my deep gratitude towards my parents and family for providing me with consistent support and continuous encouragement throughout my years of study and through the completion of researching and writing this thesis. This achievement would not have been conceivable without them. Also, I thank all those whose names, though not mentioned for their support and encouragement in completing this project. Thank you.

ABSTRACT

This study shows the automated highlight generation through the audiovisual framework for e-sport Dota 2.

Currently, in most of the video sharing platforms there are plenty of content creators on top of e-sports. Those contents are mostly based on the games played by the players. Community of the e-sport tent to watch those content through those platforms. With the high availability of contents and lack of time community missed most of the interesting videos. While discuss with the community can see that they tend to watch the highlights which is shorter version of full game play. While surfing video sharing platforms such as YouTube, stats shows that there are more views for highlights videos than full videos.

Generation of highlights of the video consists of creating a shorter version of the video which consists of the most interesting parts of the original. It requires domain knowledge and time-consuming process. In order to reduce those effort this research will facilitate efficient highlight generation tool with highlight detection and give final output as MP4 video output. This research work mainly based on the audio intensity detection, template matching to detect key events and event detection using predefined text appearing on the screen. In this research jointly uses visual features and audio features to construct highlight generation model and enable compact highlight representation.

In this highlight generation process, first identified the clips by analyzing the energy of the audio file. Identified clips passed into visual processing module. In the visual processing module that uses template matching and OCR techniques to identify the highlighted clips. Finally, all identified clips have been merged and created output in video file format.

Table of Contents

LIST OF FIGURES

LIST OF ABBREVIATION

2D : Two Dimensional
CNN : Convolutional Neural Network
OCR : Optical Character Recognition
OS : Operating System
OS X : Apple MacIntosh Operating System
RNN : Recurrent Neural Network
SVM : Support Vector Machine
UI : User Interface

1. INTRODUCTION

Production of multimedia content based on video games has been rapidly increased in recent years. Millions of videos and image content publish on the internet daily and share among the community. The term esports describes video games that are played competitively and usually watched by large audiences(Hamari and Sjöblom, 2017). Esports has become an important research field across academia and industry(Mahlmann et al., 2016) due to the availability of high-dimensional, high-volume data from virtually every match.

1.1. MOTIVATION

Dota 2 is a multiplayer online battle arena video game developed and published by a company called Value which is American based game development company("Dota 2," 2017). The development of Dota 2 began in 2009 by IceFrog and the game is now released for Microsoft Windows, OS X, and Linux via the digital distribution platform called Steam. Dota 2 is played between two teams called Dire and Radiant has five players in each, with each team defending their Ancient on the map. Each player in the team can pick a hero as his character per game. Since the team combination and item builds per hero differ from game to game each game is unique.

This game has a large e-sport community, with teams from over the world playing for many professional tournaments and leagues. Value conducts several main events for the community who can participate and win large prize pools. Some of the tournaments are DreamLeague, We play Animajor, Singapore major, Omega league…etc. The main event called The International tournament has the biggest prize pool for any e-sport. In 2019 the prize pool was $34,330,069("Esportsearnings," 2020) and in 2021 the prize pool is $40,018,195, announced to happen in Bucharest, Romania with 18 teams. Dota 2 has more than 11 million communities who play the game and most of them are addicted to watching the pro players' gameplays. As watching a Dota 2 game needs a lot of time, many players prefer to watch the summary of the games. In this sense, Dota fans have become more and more eager; this can be seen on YouTube(Pires and Simon, 2015) how many views got for gameplay highlights than full gameplay.

The motivation came for this research work came after playing the game and watching videos for over 7 years. When surfing on YouTube some Dota 2 gameplay videos have more than one million views per video.

1.2. STATEMENT OF THE PROBLEM

While going through video sharing platforms and steam platforms it can be seen that producing highlights of recorded videos for audio-visual content has been of great interest in the past few years. It's labor-intensive work that requires domain-specific knowledge with the process of generating any highlights. Generation of highlights of the video consists of creating a shorter version of the video which consists of the most interesting parts of the original. In Dota 2 contest viewers consider the most important events in the game such as Killstreak, multiple kills, Deaths, Tower destroys, Barracks destroy, Roshan kills, Courier kills, winning moments…etc. In this manual process, content creators use some video editing tools, video recording tools, monitoring tools, high-end computers…etc. This research will provide a model capable of automatically generating highlights focused on the online computer game Dota 2 which is helping to reduce the impact of the human factor for generating highlight videos.

1.3. AIMS AND OBJECTIVES

- To identify special moments in gameplay using Audio Analysis
- To identify special movements in gameplay using Video Analysis
- Generate highlight video of gameplay using identified special moments

1.4. PROJECT SCOPE

- High-resolution full gameplay capture needs for the process
- Video and audio processing, natural language processing, and image processing techniques will be used in this study.
- Only the English language will be considered in the study.

1.5. PROPOSED SOLUTION OVERVIEW

This thesis will describe how the system implemented using high-resolution video files as input and with aid of audio analyzing techniques and video analyzing using OpenCV, Tesseract and generate highlight video as output

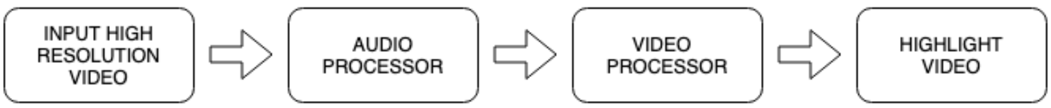


Figure 1: Overall Process

1.6. OPENCV

OpenCV stands for Open-Source Computer Vision Library. As for the definition, this is an open-source project that has more than 2500 optimized algorithms, which contains a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and distinguish faces, recognize objects, classify human actions in videos, trail camera movements, trail moving objects, extract 3D models of objects, generate 3D point clouds from stereo cameras, stitch images together to generate a high-resolution image of an whole scene, find related images from an image database, eliminate red eyes from images taken using flash, track eye movements, identify scenery and establish markers to overlay it with augmented reality(Mahamkali and Ayyasamy, 2015).

OpenCV is originally developed by using C++ programming language and it is extended to several languages including Java. Currently, these implementations support Windows, Linux, Android and Mac OS. In this research it uses OpenCV version 4.5.3.

1.7. TESSERACT

Tesseract was originally developed at Hewlett-Packard Laboratories Bristol and Hewlett-Packard Co, Greeley Colorado between 1985 and 1994, with some more changes made in 1996 to port to Windows, and some C++izing in 1998. In 2005 Tesseract was open-sourced by HP. Since 2006 it is developed by Google(Nair, 2016).

Tesseract is an open-source Optical Character Recognition (OCR) engine, which can recognize more than 100 languages and it supports unicode (UTF-8). It can be train to recognize other languages as well. This engine is used machine learning algorithms and opencv techniques to recognize characters using the training dataset. Tesseract is originally developed by using C++ and built to different language libraries such as Python, Java. Currently, Tesseract 4 is released and source code can be found at GitHub, where there is a separate repository for Tesseract(Nair, 2016).

Tesseract OCR engine internally use machine learning techniques to recognize characters from its training dataset.

1.8. GAUSSIAN FILTER

This is an image blurring technique which uses a Gaussian function. By blurring the image, it remove detail and the noise in the image for some extend. The Gaussian function of a 2D image is(H.-C. Shih, 2017),

In image processing applications, the Gaussian distribution needs to be estimated by a convolution kernel. Therefore, values from this distribution are used to form a convolution matrix then applied to the original image. Each pixel's novel value is a weighted average of that pixel's neighborhood. Thus, the original pixel's value obtains the heaviest weight (having the highest Gaussian value) and adjacent pixels receive smaller weights as their distance to the original pixel increase(H.-C. Shih, 2017).

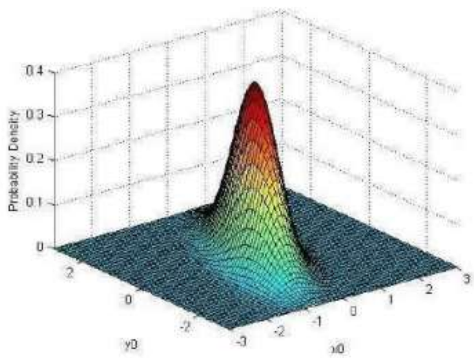


Figure 2: Gaussian distribution of a 2D Image

By following above steps original image's noise can be filtered. Gaussian filter is a low pass filter which is attenuating high frequency signals. Gaussian filter is commonly used to edge detection, medical studies.

1.9. DOTABASE

Dotabase("Dotabase," n.d.) is an Sqlite database which represents the data form the game files in Dota 2 and an sqlalchemy representation to be used with it. It's an open-source project which provide developers to access the Dota 2 files such as he sounds files, image files, hero models, scripts…etc. One main feature of this project that couldn't find anywhere else, is a representation of dota's Hero Response system. This is the system that controls the various vocal responses that heroes have to actions that are happening in the game.

1.10. STRUCTURE OF THESIS

Here, in this chapter described the Introduction of the research on, "Automated Highlights Generator for Dota 2 game Using Audio-Visual Framework". In chapter 2, it is going to discuss the Existing Approaches. After that in chapter 3, it is going to show the Methodology that have been used. In chapter 4 added the Evaluation and Results of the developed model. In chapter 5, it will give the conclusion and details about future work. At the end of the thesis added Appendices and References.

1.11. SUMMARY

This chapter provided a brief introduction to research work. It also described the aim and objectives of the proposed system. Then, it shows the used the technologies. Finally, it presented the structure of the report. The next chapter will give a critical review of similar existing systems, including the new technologies and the guidelines make our project which fit the problem.

2. LITERATURE REVIEW

2.1. INTRODUCTION

The previous chapter has provided the introduction of the research. Also, it described the background and motivation for the project and the importance of the problem. Then it described almost the aim and objectives and the proposed solution in brief. This chapter hopes to give some background information on the research. Here going to describe other approaches to the problems that were stated in the previous chapter.

2.2. TRADITIONAL SPORTS

Several studies have been conducted in the generation of the highlights videos on sports such as football(Tedesco et al., n.d.), cricket(Shukla et al., 2018)(M. H. Kolekar and S. Sengupta, 2008), soccer(Raventós, n.d.), basketball, ...etc as well as e-sports(Raventós, n.d.)(Chu and Chou, n.d.)(Song, 2016)(Mahlmann et al., 2016). Other than generating highlights using the research work has been conducted to analyze the player skills, weak points, strengths, suspicious behaviors…etc.

The research has been conducted by Anant Baijal, Jaeyoun Cho, Woojung Lee and Byeong-Seob Ko have conducted a study on Sports highlights generation based on acoustic events detection(Baijal et al., 2015). In this research, they have only derived information such as Referee Whistle, Commentator's excitement Speech, Crowd noise from the audio to generate highlight videos. In this research, multi-stage classification is used to detection of key events which act as a highlight scene. In the first stage they used acoustic events detection with high recall rate and negligible precision error. Then by using detected events generated the highlights. In this research work Labeled data used to preprocessing and feature extraction. In feature extraction employed Mel Frequency Cepstral Coefficients and their first order differential coefficients, commonly known as delta-MFCC. Gaussian Mixture Models used to learn the extracted features. In that research they have used rugby as the case study, and they have achieved 97% Try Recall and 93% Highlights Precision as the results. This system was proposed to embed in consumer electronics and use to generating highlights in online broadcasts as well as in offline scenarios such as prerecorded videos.

P. Karthivel K.P. Soman and M. Sabarimalai have conducted a study on Automated Referee Whistle Sound Detection for Extraction of Highlights from Sports Video(P et al., 2011). In that audio signal extracted separately then pre-process, filter-blank, Band pass filtering and extracted the highlighted moments. The referee whistle detector is designed for exploring the possibility to build a unified framework to identify highlights from sports videos. This proposed method divided the audio into 6-second non-overlapping blocks and each block was processed to detect the whistle sounds. This model was tested using an American Football audio stream of a 2-hour and 11-minute which includes game sounds and high background noise such as crowd cheering, applauses, and exited commentary.

The research was published in 2018 under the topic of Automatic Cricket Highlight generation using Event-Driven and Excitement-Based features(Shukla et al., 2018) based on both audio and video. In this research it has been used image segmentation for detecting key events, Optical Character Recognition (OCR) engine for score detection, A CNN+SVM methodology to classify the keyframes, Audio intensity to detect crowd excitement. In this study the given video input has been segmented in early stages. The replays are detected by extracting frames from those segments. The cropped image from the frame is then used for classifying a video shot as an on-going event of the game or as a replay/advertisement using a Convolutional Neural Network (CNN) + Support Vector Machine (SVM) framework. This framework relies on 4096 representations extracted from the fc7 layer of the AlexNet network, which are fed to a trained Linear Support Vector Machine that has been trained on representations of 2000 images of both classes. In this study showed the framework capable of achieving comparable results to manual highlights and that it yields acceptable results for cricket fans.

P. Chappidi, K.R. Kothinti, and K.R. Namuduri has conducted study on Automatic Extraction of Highlights from a Cricket Video using HMM and MPEG-7 Descriptors(Namuduri, 2009). In this study proposed Hidden Markov Model (HMM) to model the highlight. Every shot and its transition to the next shot are compared with the model. The shots that have same transition pattern as that of the HMM are extracted to form the highlight. In this study used the color variations of the different areas of the ground such as grass color, pitch color, texture, motion and number of edges in the frames.

2.3. E-SPORTS

With the advent of live streaming systems such as Twitch, Youtube, Facebook, GoodGame, many works have been proposed on top of those systems from various perspectives. Kaytoue et al(Kaytoue et al., 2012) focused on electronic sports videos streamed by Twitch and advocated that much potential revenue can be made to professional gamers, casters, and streaming platforms. They have proposed a model to predict the number of viewers and category of viewers for a specific Twitch channel. There are a few numbers of research were conducted for game videos with the perspective of audio and visual analysis. Several image processing and computer vision techniques have been utilized by Douglass to gameplay recordings. Lewis et al.(Lewis et al., n.d.) analysed player's actions, such as actions per time frame and spatial variance of action, to discover the correlation between actions and winning games. Not surprisingly, they found that gamers able to most quickly execute actions tend to win. Rioult et al.(Rioult et al., 2014) extracted topological clues, such as the area of a polygon where players move and the inertia of the team, to predict outcomes of multi-player online battle arena games.

Wei-Ta Chu and Yung-CheinChu have conducted a study on Event Detection and Highlight Detection of Broadcasted Game Videos(Chu and Chou, n.d.) which is focused on game event detection, highlight detection by recognizing pre-defined text on the screen. In that paper, they have used the League of Legends game as a cause study and used the Tesseract OCR package to recognize the predefined text which will appear in the key events. This study used Arousal Model that the level of arousal of a user rises as a consequence of the increase of various stimuli, in the representation of various features in this work. This study achieved 92% average event detection accuracy. This study has been evaluated by 24 games of League of Legends World Cham- pion broadcasted by Twitch games in 2014, which can be grouped into six game series, and each series comprises multiple games between two teams.

With the spread of eSports championships, a video-based highlight encounter detector for Multiplayer Online Battle Arena (MOBA) games was proposed(Song, 2016). The author proposed various solutions for frame-wise classifiers based on CNN and RNN, considering both single and cascaded architectures, and different shapes for the output; in fact, the data set was tagged considering four different levels of highlight, starting from non-highlight up to maximum relevance. The peak performances were achieved, mostly, considering only a binary output: one of the considered games reported precision of 83.2% and a recall of 86.3%. A point to stress out about this model is that it was designed to work with real-time video streams.

Research has been conducted on Event Detection and Highlight Detection of Broadcasted Game Videos(Chu and Chou, n.d.) based on the League of Legends which is similar to Dota 2. It has been used predefined text appears on screen to detect highlights and events. In this research they employed Tesseract OCR package to recognise the text appearing on the screen. Through recognizing predefined text displayed on screen, they detected events to ease direct access. For highlight detection, they described visual appearance, events, and viewer's reaction, and then construct two highlight models. Evaluation on famous game videos shows that the proposed methods yield accurate event detection and promising highlight extraction performance.

2.4. SUMMERY

This literature review revealed that Neural network, Image processing technologies, OCR, Sound event classification has been widely used in the video summarization field. According to the literature, it can be improving the accuracy when the data is pre-processed by removing noise. The next chapter will explain the methodology in this research work.

3. METHODOLOGY

3.1. INTRODUCTION

In the previous chapter discussed about the technologies that adapt to the research. This chapter mainly describes about the design and implementation to the solution made, gives a description about which inputs have used and inputs, outputs and process. It also includes the main structure of the system.

There are several important phases in the Dota 2 game video to identify before processing.

- Cinematic pre-casting
- Hero drafting phase
- Game play phase
- End game phase

Figure 3: Cinematic casting view

Figure 4: Hero drafting phase

Figure 5: Game Play Phase

Figure 6: End game phase

These phases occur in game as sequence above mentioned order. Before starting the game play part of the video can be labelled as drafting phase. Game play part can be labelled using special characteristics of the UI in game and End game part can be labelled using victory screen.

There are several important areas in game play screen to identify before understanding the highlight generation process. Those are,

i. Event Banner
ii. Event indicator

iii. Map

iv. Player chats

v. Hero Bar

vi. Player/Hero stats

vii. Selected player

viii. Selected player current gold

Figure 7: Sample image of gameplay

When special events occur, special effects often appear such as kill banners, hero skills, text changing…etc.

Table 1: Some Special event text

Text	Description	Major/ Minor
{P_x} drew first blood by killing {P_y B}	First kill in game	Major
{P_x} {kill icon} {P_x}	Side panel indicate every kill	All Events
{P_x} bounty rune 40 each	Collect bounty rune	Minor
Radiant structured fortified	Fortification activated	Minor
Dire structured fortified		Minor
{P_x} destroy tower	Tower destroyed	Minor
{P_x} is on a killing spree	Player has 3 kills steak	Major

{P_x} is Dominating	Player has 4 kill steak	Major
{P_x} is on a Mega kill	Player has 5 kill steak	Major
{P_x} is Unstoppable	Player has 6 kill steak	Major
{P_x} is Wicked sick	Player has 7 kill steak	Major
{P_x} is on a Monster kill	Player has 8 kill steak	Major
{P_x} is on a Godlike	Player has 9 kill steak	Major
{P_x} is on Beyond Godlike	Player has 10 or more kill steak	Major
{P_x} is unstoppable with a double kill	Get two kills in short period	Major
{P_x} is killing spree with a triple kill	Get three kills in short period	Major
{P_x} is mega kill steak with a ultra-kill	Get four kills in short period	Major
{P_x} is unstoppable with a rampage	Get five kills in short period	Major
{T_x} VICTORY	Team has won the game	Major

3.2. OVERALL SYSTEM DESIGN

For extract highlight in game play video this research will extract features from both video frames and audio. Constructed model using both features will be used to construct the model. This proposed model contains two separate modules related to audio processing and video processing.

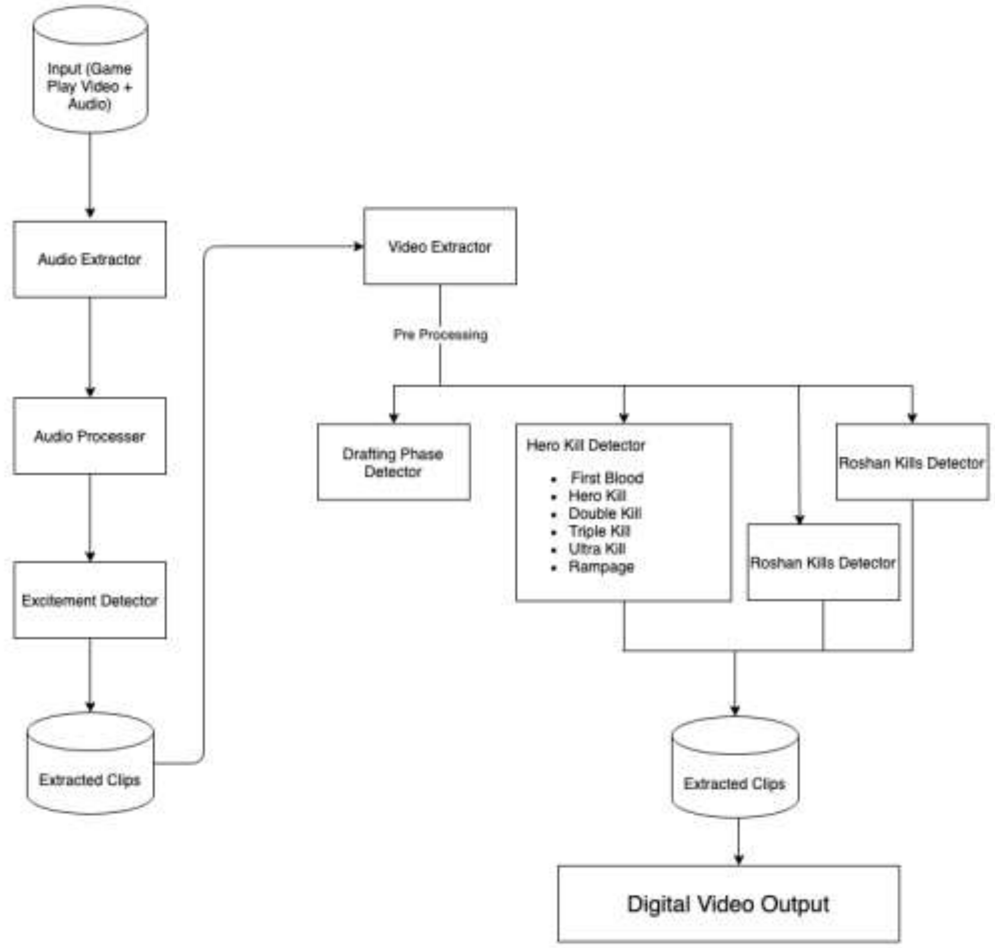


Figure 8: Proposed System Design

3.3. AUDIO PROCESSING

In the system will extracted audio file as wav format and it will be converted to mono channel by averaging both channel data. In audio that consists of caster audio, crowd cheering and the game audio. Game audio comes with different announcer pack which indicates special events mentioned in Table 1: Some Special event text and hero skills sound effects for different powers. Combination of more hero skill sounds indicate that heroes battling with each other's. When casting game by a casters level of caster voice is changing time to time according to the situation of the game. Histogram of Gradients of Time-Frequency will be used to detect audio scene.

For detecting the exactment of the crowd or the casters will use the intensity level of the created mono wav file. Louder the sound generates higher amplitude and lower the sound generate lower amplitude because the intensity of sound is proportional to the square of amplitude. Intensity is defined to be the power per unit area carried by a wave. The intensity of a sound depends upon its pressure amplitude. The relationship between the intensity of a sound wave and its pressure amplitude (or pressure variation Δp) is where P is the power through an area A,

$$I = \frac{P}{A}, \qquad I = \frac{(\Delta p)^2}{2\rho v_w},$$

Figure 9: Equations to calculate intensity levels

By applying moving average filtering, it will reduce the noise from the audio signal. Using that moving averaged results that will calculate the threshold values for particular section of audio to detect excitement timelines. Here use the down sampling techniques to improve the memory efficient because the input video length is normally around 1 hour. In below graph horizontal line indicates the dynamically created threshold value for detection excitements.

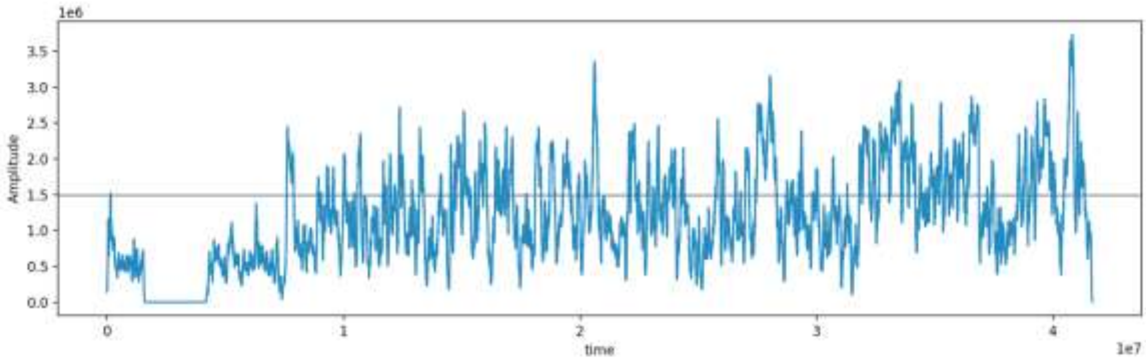


Figure 10: Amplitude vs Time plot for audio file

Because of some instance noise there are some clips which are in small time frames. To avoid that those noise clips only selected the clips which have considerable time. The derived time frame will pass to next stage for video processing.

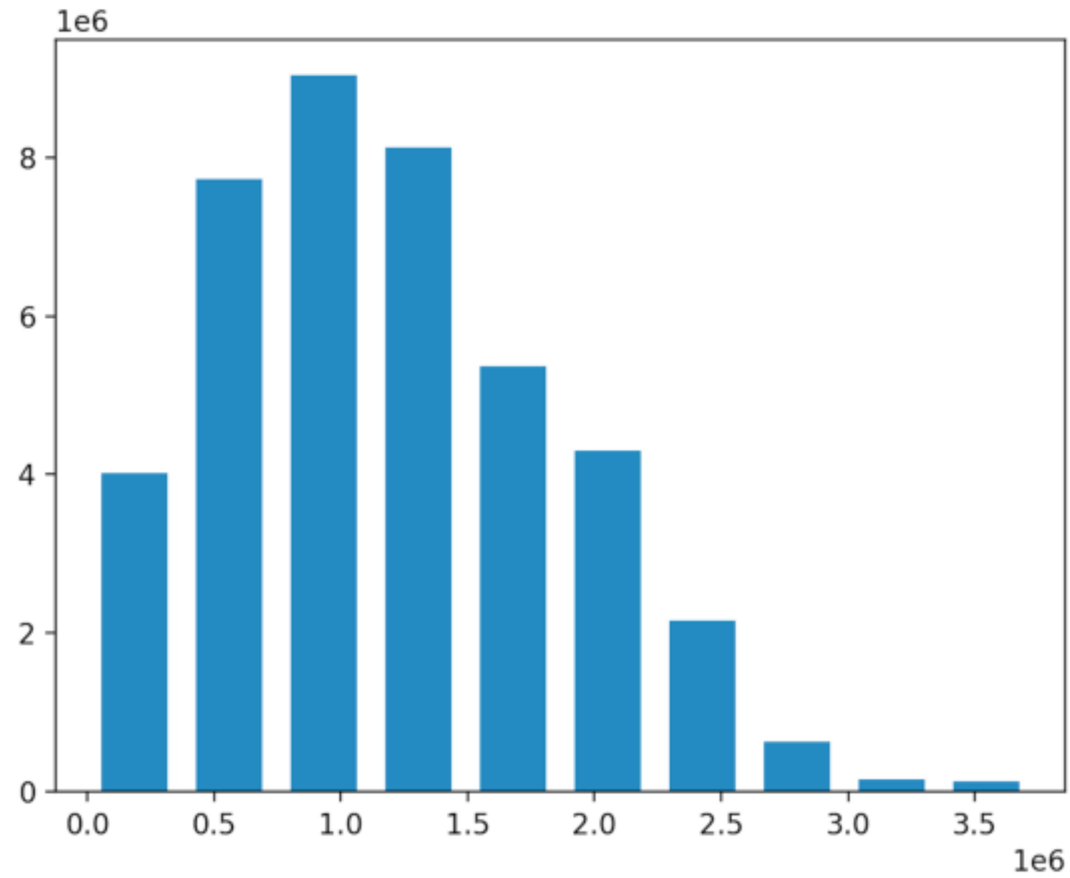


Figure 11: Short Time Energy Distribution

3.4. VIDEO PROCESSING

In this Video processing module use, given raw video file and timeline which is derived from the Audio processing module as the inputs. Since the Audio processing module detects the excitements, that timeline may contain unnecessary clips which can be omitted in the highlight video. If all the clips derived from the above model, output will be lengthier. In this module can fine tune the output to the acceptable accuracy.

Video processing module contains three main parts,

- Game Start identifier
- Key Event Detector

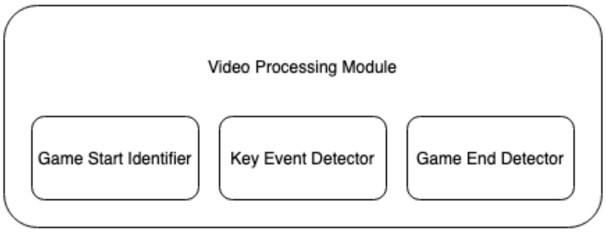


Figure 12: Video processing module

As the first step of this module get 2 frames per second in time bound of above mentioned. Those frames are used to above modules in video processing. The system reads the image by converting the input image to a grayscale image.

3.4.1. GAME START IDENTIFIER

When game hero drafting phase begins "Hero Bar" does not appear and it is starting to appear when game starts. This will be derived by calculating the defined width height relative to the resolution of the derived frame. In this Hero Bar contains 10 heroes battling in the game, game timer, day/night indicator and the kills each team got up to the current state of the game.

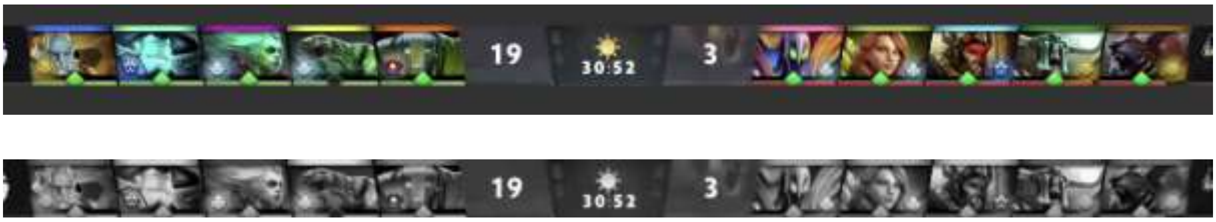


Figure 13: Hero Bar Detected Colored Vs Grayscale

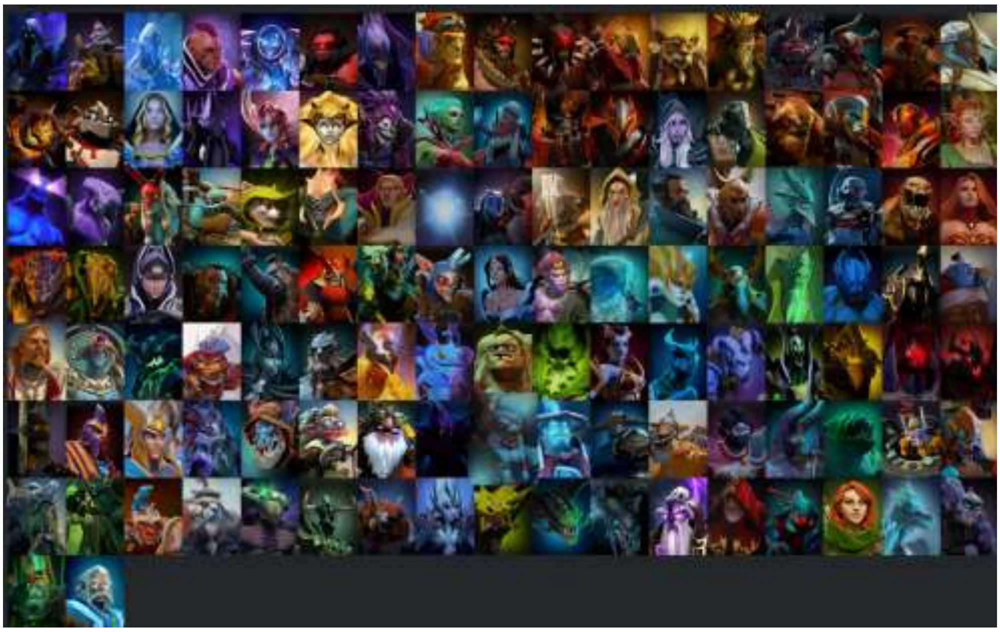


Figure 14: Dotabase Hero Templates

In this derived hero bar will match using template matching algorithms with hero templates which will be obtained by Database resources. The game starting time detect if the 10 heroes appearing in the Hero Bar with more than 0.8 confidence. So, the other clips before to that time will be ignored and that will not be considered as part of highlights even though that was detected as excitement events. For the Figure 13, the implemented model has detected the heroes correctly with the following score.

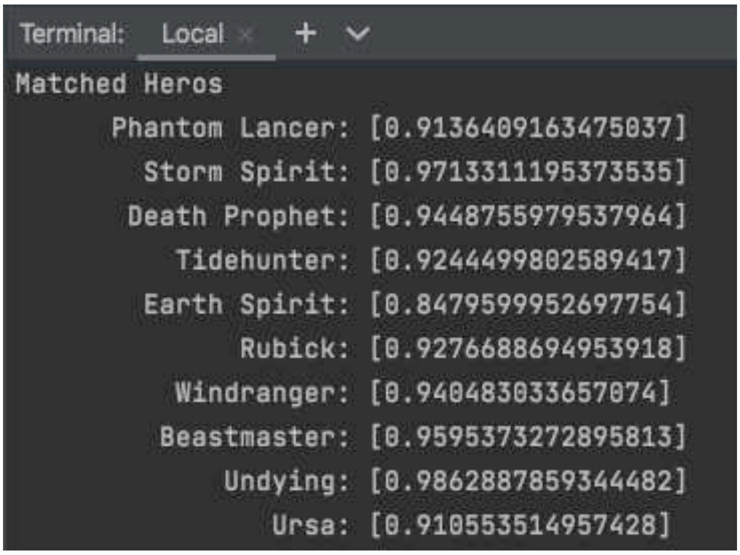


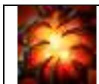
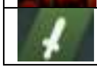
Figure 15: Console Output Hero List Detected with Confidence

3.4.2. KEY EVENT DETECTION

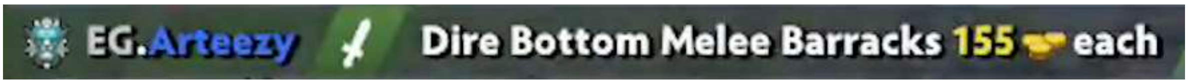
The events happen during the game play can be categorized into major and minor events subject to highlights generation. In highlights no need to include each and every events happen in the game play. For this study it has been selected few major events as follows.

- Destroy the towers
- Hero Kills (Double Kill, Triple Kill, Ultra Kill, Rampage)
- Roshan Kills
- Scan by a team
- Victory

For the Hero Kill, Roshan Kills, Scan by a team there is indicators appearing on the Event Indication section. Those events have been identified using template matching. If any clip includes the special events that has been added to final highlights.

	Roshan Kill template
	Hero Kill/Tower destroy template

	Scan in progress template



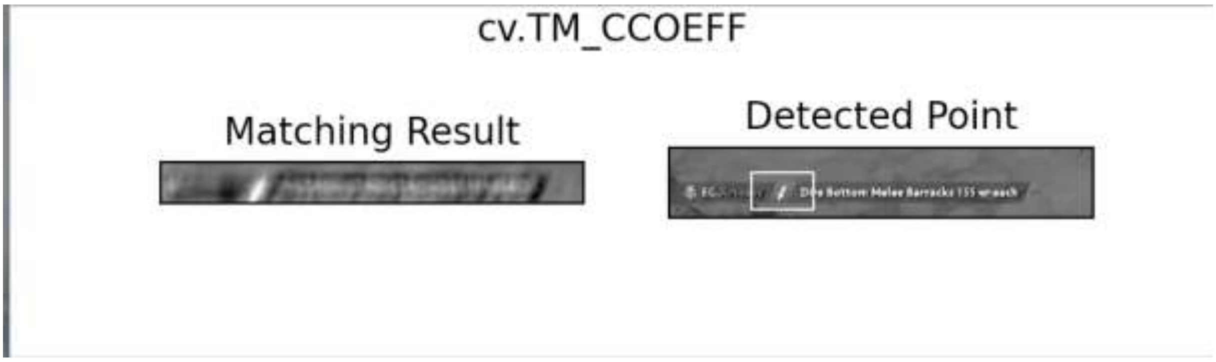


Figure 16: Special event detection using Template matching

For special events such as Victory, Double Kill, Triple Kill, Ultra Kill…etc text banner will appear on Event Banner section. Those events are identified by processing the kill banner area in the derived frame. From the frame system will crop the kill banner area by calculation the size of area relative to the resolution of the frame.

Figure 17: Cropped Kill Banner Area

This detected frame will convert into gray scale image and apply adaptive thresholds algorithm for that image. The detected area will apply the canny edge detection algorithm before pass that into Tesseract OCR.

Figure 18: Canny edge detected selected area

The output from the OCR process will be checked with the pre-defined texts which is appearing on the Kill Banner to determine whether it is special event or not.

4. EVALUATION AND RESULTS

4.1. ITRODUCTION

The above chapter described about the design and implementation of the proposed solution. This chapter includes a evaluation results about the things in this research.

4.2. EVALUATION CRITERIAS

When planning performance evaluation of an image processing application, there could be situation that an algorithm processes an image with rare or unusual features. Such situations, the performance depends on the performance indicators. Typical performance indicators include [1]

1. Accuracy: algorithm's performance with respect to some reference
2. Robustness: algorithm's ability for enduring numerous conditions
3. Sensitivity: algorithm's responsive to small changes in structures
4. Adaptability: algorithm's behavior with variation in images
5. Reliability: output of the algorithm when repeatedly using the same stable data
6. Efficiency: the practicability of an algorithm (space and time)

In this study, it is considering image processing domain area with template matching, optical character recognition which is printed in game screen and audio processing area with intensity based excitement detection.

Canny Edge Detection used since its efficiency is higher, it is well fit with simple images and complex images as well and higher signal to noise ratio than the other edge detection(Chandwadkar et al., 2013)

Parameters	Sobel	Canny
Computation[5]	Simple and time efficient	Complex and time consuming
Signal to Noise ratio	Low	High
Texture based image Fig.2	Less efficient	More efficient
No of objects in image Fig.3, Fig.4	Suitable for simple images	Suitable for simple as well as complex images
Application area /domain [5] [6]	Massive data communication and data transfer	Medical field for X-ray diagnosis and object recognition

Figure 19: Comparison between Sobel Edge Detection and Canny Edge Detection(Chandwadkar et al., 2013)

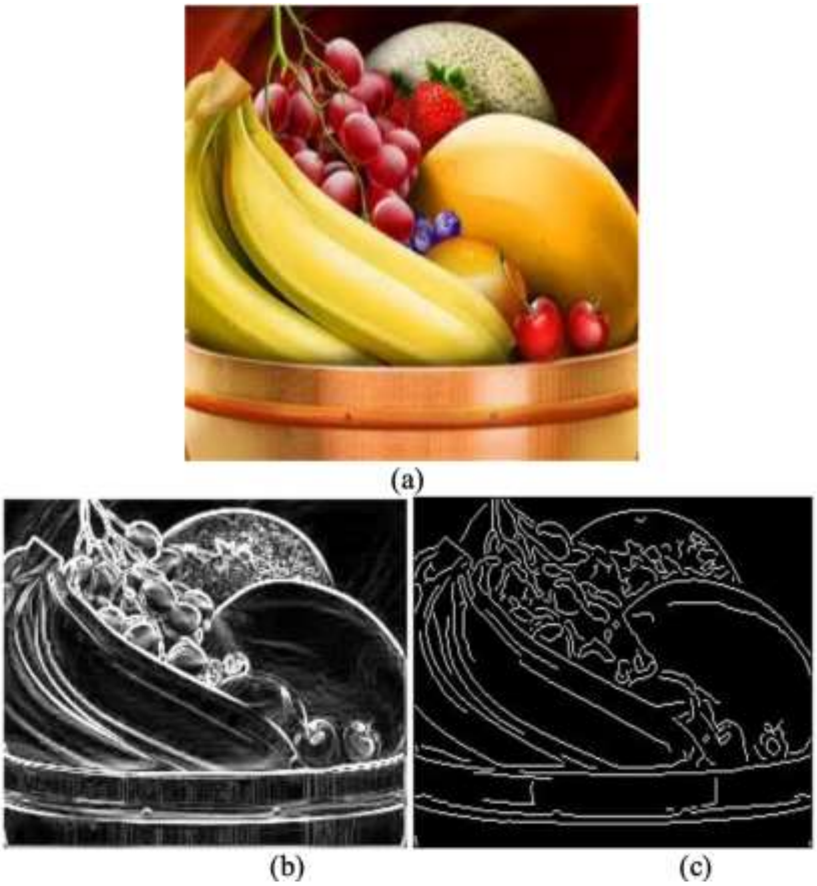


Figure 20: Comparison between SobelAEdge Detection and CannyAEdge Detection (a) original Image (b) Sobel output (c) canny output(Chandwadkar et al., 2013)

To identify Kill Banner data, it is used OCR technique to extract the text from the frames. Tesseract is used as the OCR because currently it is a better and free open-source OCR engine compared to GOCR and Transym(Patel et al., n.d.). Accuracy of identifying different types of fonts, different resolution levels, different brightness levels and performance speed is better in Tesseract with compared to other open-source OCR engines. Comparison between Tesseract and GOCR is as follows,

BRIGHTNESS	TOTAL CHARACTERS EXTRACTED BY TESSERACT	CHARACTERS CORRECTLY EXTRACTED BY TESSERACT	TOTAL CHARACTERS EXTRACTED BY GOCR	CHARACTERS CORRECTLY EXTRACTED BY GOCR	TESSERACT ACCURACY (IN %)	GOCR ACCURACY (IN %)	TESSERACT PRECISION (IN %)	GOCR PRECISION (IN %)
25	39	37	28	23	94.8	58.9	94.8	82.1
50	39	38	27	26	97.4	66.6	97.4	96.2
100	39	37	1	1	94.8	02.5	94.8	100

Figure 21: Comparison between Tools using different Brightness values(Patel et al., n.d.)

IMAGE TYPE	TOTAL CHARACTERS EXTRACTED BY TESSERACT	CHARACTERS CORRECTLY EXTRACTED BY TESSERACT	TOTAL CHARACTERS EXTRACTED BY GOCR	CHARACTERS CORRECTLY EXTRACTED BY GOCR	TESSERACT ACCURACY (IN %)	GOCR ACCURACY (IN %)	TESSERACT PRECISION (IN %)	GOCR PRECISION (IN %)
Color	39	38	28	25	97.4	64.1	97.4	89.2
Gray scale	39	38	24	18	97.4	46.1	97.4	75.0
Black and White	39	38	27	19	97.4	48.7	97.4	70.3

Figure 22: Comparison between Tools using different Image types(Patel et al., n.d.)

IMAGE TYPE	RESOLUTION	TOTAL CHARACTERS EXTRACTED BY TESSERACT	CHARACTERS CORRECTLY EXTRACTED BY TESSERACT	TOTAL CHARACTERS EXTRACTED BY GOCR	CHARACTERS CORRECTLY EXTRACTED BY GOCR	TESSERACT ACCURACY (IN %)	GOCR ACCURACY (IN %)	TESSERACT PRECISION (IN %)	GOCR PRECISION (IN %)
Color	75	39	38	39	38	97.4	97.4	97.4	97.4
Color	300	39	38	_	_	97.4	_	97.4	_
Color	1200	0	0	_	_	_	_	_	_
Gray scale	75	_	_	39	38	_	97.4	_	97.4
Gray scale	300	_	_	_	_	_	_	_	_
Gray scale	600	39	38	4	2	97.4	05.1	97.4	50.0
Black and White	75	39	38	39	38	97.4	97.4	97.4	97.4
Black and White	300	39	38	_	_	97.4	_	97.4	_
Black and White	1200	_	_	_	_	_	_	_	_

Figure 23: Comparison between Tools using different Resolution values(Patel et al., n.d.)

This research used template matching for identify heros selected playing in the game and match with templates such as scan, kill, Roshan…etc to detect the key evets occurring in the game play. Following tables illustrates the accuracy of the different template matching algorithms.

Method	Multi-view dataset		Category dataset	
	$E_{95\%}$,%	EER,%	$E_{95\%}$,%	EER,%
PQ-HOG, *t*=4	30.6 (*)	12.4 (*)	48.2	17.3
BiCE PCA	12.76(#)	7.03(#)	45.04	16.74
SIFT 64	33.38(#)	11.51(#)	87.37	32.53

Figure 24: Comparison between template matching algorithms(Sibiryakov, 2011)

		Faces	Cars	Motorbikes
Number of images		450	126	826
Mean image size, pix.		896×592	896×592	554×368
Template size, pix.		240×234	466×344	162×190
PQ-HOG parameters:	N_g pix.	8	10	10
	N_{bits}	8	24	24
	t, bits	3	3	3
Correctly matched, %	PQ-HOG	**95.5**	**98.4**	**61.1**
	NCC	89.5	92.7	33.5
Mean run time per image (ms.)	PQ-HOG	**52.1**	**266.5**	**30.2**
	NCC	584.8	658.9	253.1

Figure 25: Performance of PQ-HOG and NCC in three dataset(Sibiryakov, 2011)

4.3. RESULTS

In this developed model is tested with 10 full game play videos in 1080p resolution which collected from the YouTube. System has monitored in two steps and collected results as Audio Analyzing module and Final Output. Following table shows the output from the audio analyzing module.

Game Details			Audio Analyzer				
Original Video Duration	Roshan Kills	Total Kills	Detected Kills	Missed Kills	Kill Accuracy	Output Duration	Compression Ratio
3779	4	60	58	2	96.66666667	1084	3.486162362
2777	1	51	50	1	98.03921569	844	3.29028436
2185	1	47	44	3	93.61702128	700	3.121428571
2581	1	32	31	1	96.875	717	3.59972106
3402	2	35	33	2	94.28571429	815	4.174233129
2706	2	35	33	2	94.28571429	702	3.854700855
3010	2	39	37	2	94.87179487	814	3.697788698
3402	0	34	29	5	85.29411765	412	8.257281553
2739	3	56	51	5	91.07142857	835	3.280239521
3387	2	39	37	2	94.87179487	807	4.197026022
					93.98784682	773	4.095886613

Figure 26: Stats related to audio processing module

Above table shows the detection accuracy of selected special events. In overall the audio processing module achieved kill detection accuracy 93.9% and compression ratio for original video to highlight 4.09.

Game Details	Combination of Audio Analyzer and Video Frame Analyzer	
Original Video Duration	Output Duration	Compression Ratio
3779	1057	3.575212867
2777	844	3.29028436
2185	686	3.185131195
2581	661	3.904689864
3402	798	4.263157895
2706	678	3.991150442
3010	764	3.939790576
3402	389	8.745501285
2739	806	3.398263027
3387	779	4.3478819
	746.2	4.264106341

Figure 27: Stats for outcome

Above table shows the final results for the evaluated data set. In the final output the system achieved compression ratio 4.26. The system gave the highlights video file with the average length of. 12min 26sec(746 seconds).

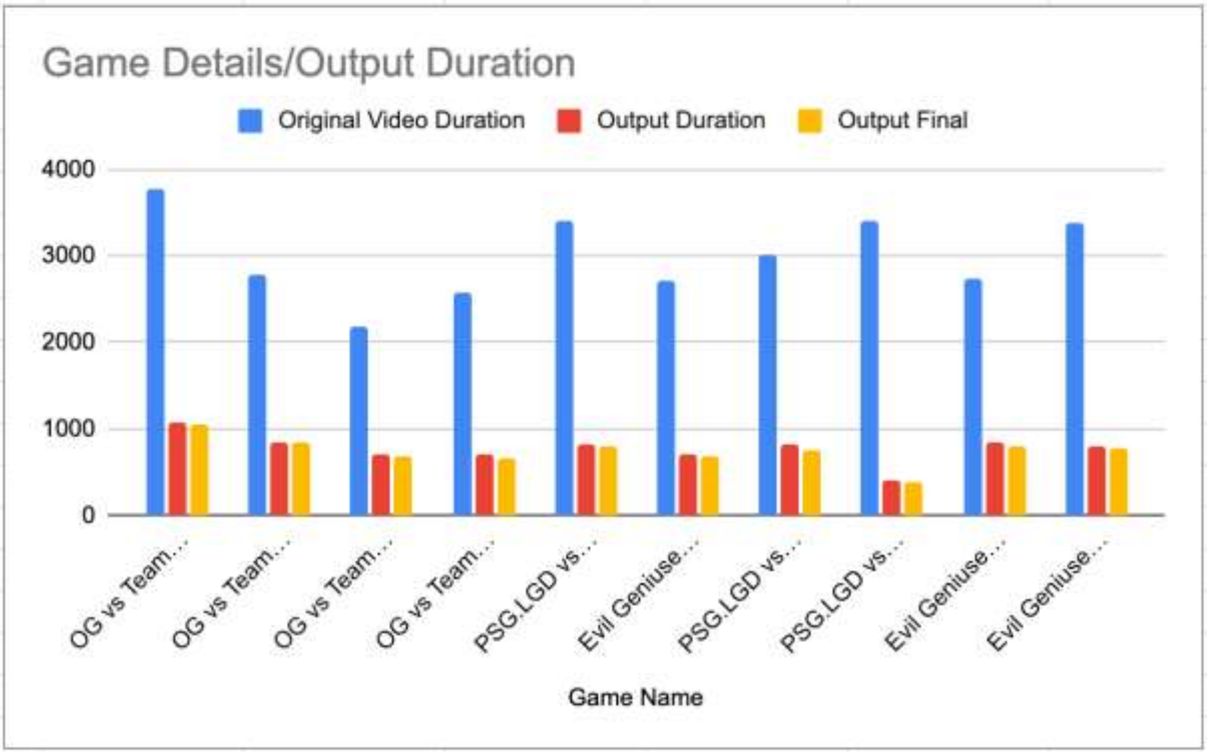


Figure 28: Input file length to Highlight Length

Following results have been obtained by the system for the evaluated videos,

- Average kill detection Accuracy: 93.99%
- Average length for highlight using audio analyzer: 773 Seconds (12min 53Sec)
- Average video Compression ratio using audio module: 4.09
- Average highlight length for final output: 746 (12min 26sec)
- Average video compression ratio for final output: 4.26

The implemented system's output compared with highlights published in the YouTube. Most of the tournament games are published in the YouTube official channels in the format of full game play and highlight game play which is created by content creators. Those data collected by manually from the YouTube and compared with the output. The results show 82 seconds standard deviation for differences between output length of proposed system and the already published videos on the YouTube. Following table and graph show the compression of the system final output with the official highlights.

	Game Details	Combination of Audio Ana	Youutube highlight details	
Game Name	Original Video Duration	Output Duration	URL	Duration
OG vs Team Liquid _	3779	1057	https://www.youtube.com/	950
OG vs Team Liquid _	2777	844	https://www.youtube.com/	649
OG vs Team Liquid _	2185	686	https://www.youtube.com/	643
OG vs Team Liquid _	2581	661	https://www.youtube.com/	692
PSG.LGD vs Virtus F	3402	798	https://www.youtube.com/	780
Evil Geniuses vs OG	2706	678	https://www.youtube.com/	745
PSG.LGD vs Team L	3010	764	https://www.youtube.com/	783
PSG.LGD vs Virtus F	3402	389	https://www.youtube.com/	780
Evil Geniuses vs OG	2739	806	https://www.youtube.com/	760
Evil Geniuses vs OG	3387	779	https://www.youtube.com/	788

Figure 29: Comparison YouTube highlights with output by system

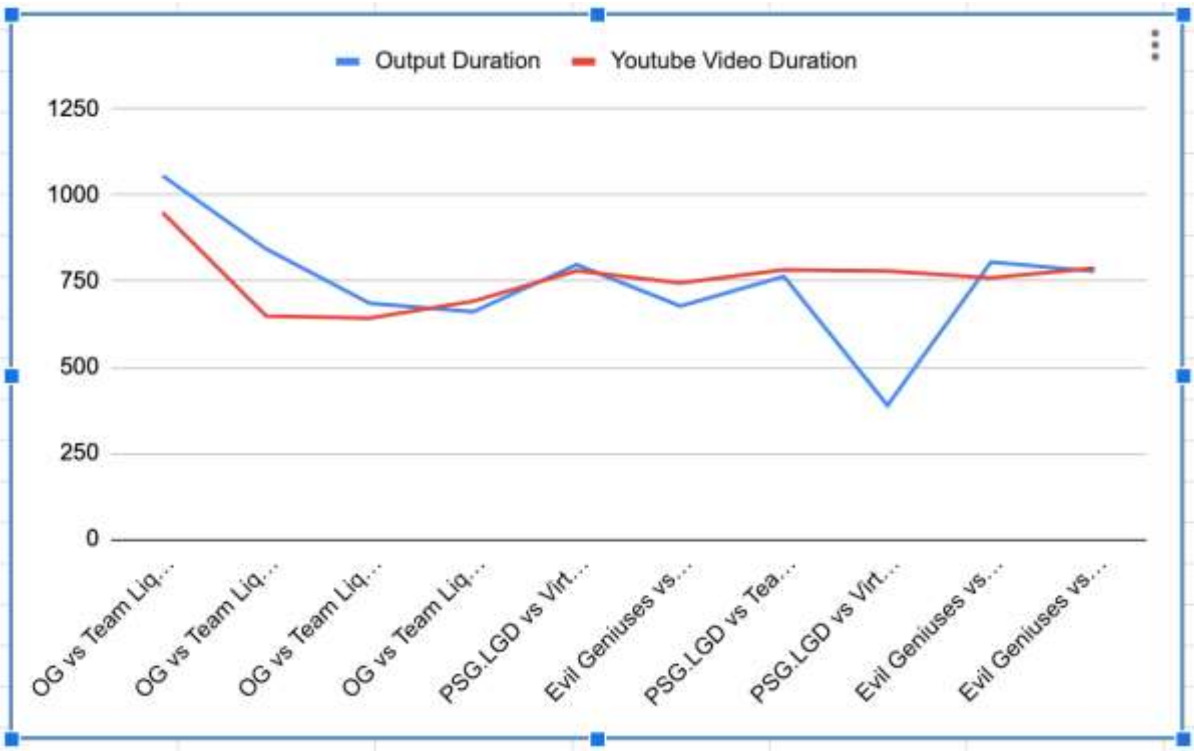


Figure 30: Graph YouTube highlight and output developed system

5. CONCLUSION & FUTURE WORK

5.1. INTRODUCTION

In the previous chapter have discussed the evaluation methodologies and demonstrate the evaluation results for this research. This chapter provides a summary and further work.

5.2. CONCLUSION

In this study, it is implemented a system to generate highlights process by using python programming language, audio processing and image processing. Identified excitements clips have been used by the video analysis. When plating the game, the given frame can be segmented as Hero Bar, Key events indicator, Kill Banner, Selected Hero, Mini Map…etc. Those regions pre-processed before applying the template matching and OCR techniques. By combining these methodologies final output generated as video file. This research work will fill the lack of highlight generator for Dota 2 game.

5.3. FUTURE WORK

The output length of the current system cannot be controlled by the user and the system is restricted for the English language.

As for the future work, this system can enhance to generate highlights with the weighted event detection which help to give score to the highlight clips. Then user will be able to trim the final video output as they prefer. Currently this system supports only for the English language. But the game is supporting multi languages. This can be extended to support multi languages. Furthermore, the audio analyzing module can be extended by adding audio classification algorithms with game sounds which are available through the Dotabase API.

6. REFERENCES

Baijal, A., Cho, J., Lee, W., Ko, B.-S., 2015. Sports highlights generation based on acoustic events detection: A rugby case study. 2015 IEEE Int. Conf. Consum. Electron. ICCE 20–23. https://doi.org/10.1109/ICCE.2015.7066303

Chandwadkar, R., Dhole, S., Gadewar, V., Raut, D., Tiwaskar, S.A., 2013. COMPARISON OF EDGE DETECTION TECHNIQUES 4.

Chu, W.-T., Chou, Y.-C., n.d. Event Detection and Highlight Detection of Broadcasted Game Videos 8.

Dota 2 [WWW Document], 2017. URL http://www.dota2.com/play

Dotabase [WWW Document], n.d. URL http://dotabase.dillerm.io/ (accessed 9.8.21).

Esportsearnings [WWW Document], 2020. URL https://www.esportsearnings.com/tournaments

Hamari, J., Sjöblom, M., 2017. What is eSports and why do people watch it? Internet Res. https://doi.org/10.1108/IntR-04-2016-0085

H.-C. Shih, 2017. A survey on content-aware video analysis for sports. Presented at the IEEE Transactions on Circuits and Systems for Video Technology.

Kaytoue, M., Silva, A., Cerf, L., Meira Jr, W., Raïssi, C., 2012. Watch me playing, I am a professional: a first study on video game live streaming. WWW12 - Proc. 21st Annu. Conf. World Wide Web Companion. https://doi.org/10.1145/2187980.2188259

Lewis, J.M., Trinh, P., Kirsh, D., n.d. A Corpus Analysis of Strategy Video Game Play in Starcraft: Brood War 6.

M. H. Kolekar, S. Sengupta, 2008. Caption content analysis based automated cricket highlight generatio. Presented at the National Communications Conference (NCC), Mumbai, pp. 461–465.

Mahamkali, N., Ayyasamy, V., 2015. OpenCV for Computer Vision Applications.

Mahlmann, T., Schubert, M., Drachen, A., 2016. Esports Analytics Through Encounter Detection.

Nair, A., 2016. Overview of Tesseract OCR engine.

Namuduri, K., 2009. Automatic extraction of highlights from a cricket video using MPEG-7 descriptors. Presented at the 1st International Conference on Communication Systems and Networks and Workshops, COMSNETS 2009, pp. 1–3. https://doi.org/10.1109/COMSNETS.2009.4808848

P, K., Manikandan, M., Kp, S., 2011. Automated Referee Whistle Sound Detection for Extraction of Highlights from Sports Video. Int. J. Comput. Appl. 12. https://doi.org/10.5120/1729-2340

Patel, C., Mohanbhai, S.C., Mohanbhai, S.C., Patel, D., Mohanbhai, S.C., n.d. Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study.

Pires, K., Simon, G., 2015. YouTube Live and Twitch: A Tour of User-Generated Live Streaming Systems. Proc. 6th ACM Multimed. Syst. Conf. MMSys 2015. https://doi.org/10.1145/2713168.2713195

Raventós, A., n.d. Audiovisual framework for automatic soccer highlights generation 78.

Rioult, F., Métivier, J.-P., Helleu, B., Scelles, N., Durand, C., 2014. Mining Tracks of Competitive Video Games. AASRI Procedia 8, 82–87. https://doi.org/10.1016/j.aasri.2014.08.014

Shukla, P., Sadana, H., Bansal, A., Verma, D., Elmadjian, C., Raman, B., Turk, M., 2018. Automatic Cricket Highlight Generation Using Event-Driven and Excitement-Based Features, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Salt Lake City, UT, USA, pp. 1881–18818. https://doi.org/10.1109/CVPRW.2018.00233

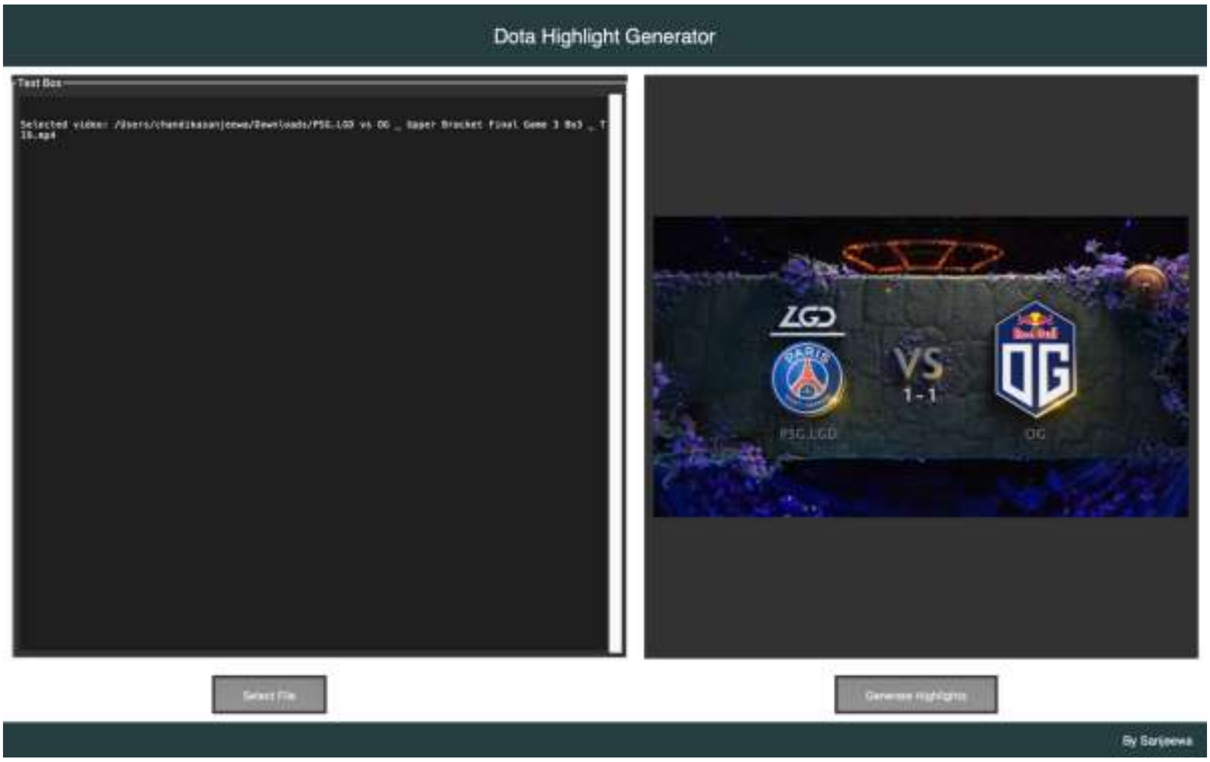
Sibiryakov, A., 2011. Fast and high-performance template matching method. Presented at the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1417–1424. https://doi.org/10.1109/CVPR.2011.5995391

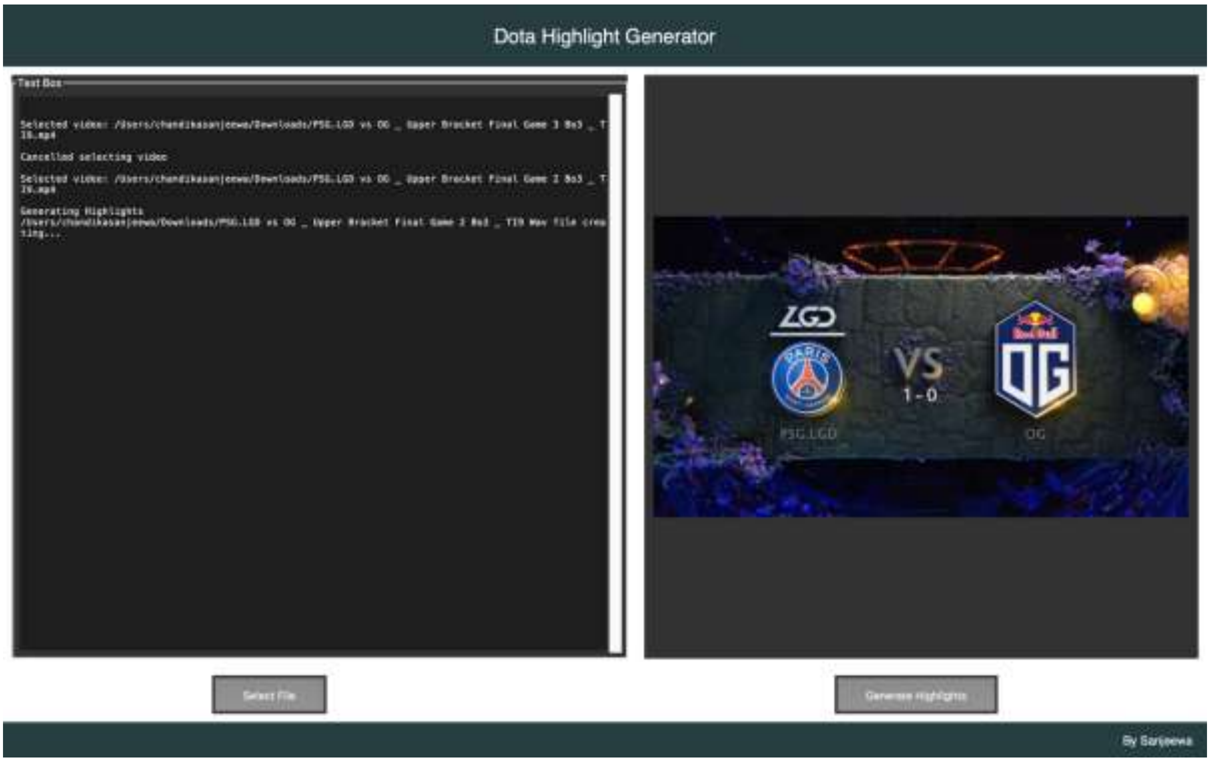
Song, Y., 2016. Real-Time Video Highlights for Yahoo Esports. ArXiv.

Tedesco, R., Sbattella, L., Scotti, V., n.d. SFERAnet: automatic generation of football highlights. Proocedings He 6th Int. Conf. Comput. Sci. Eng. Inf. Tecnol.

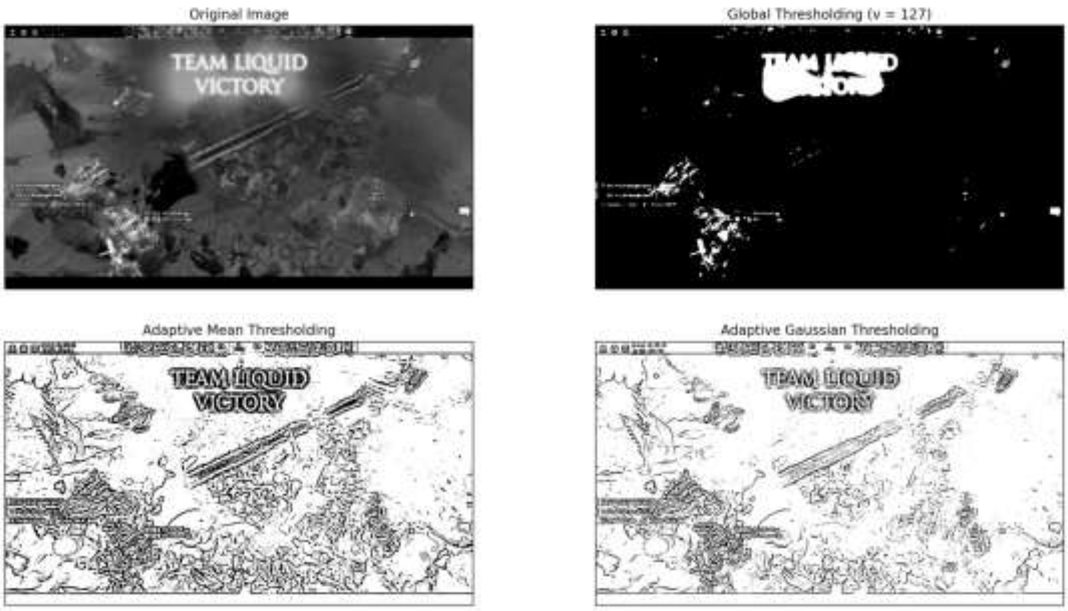
Appendix A

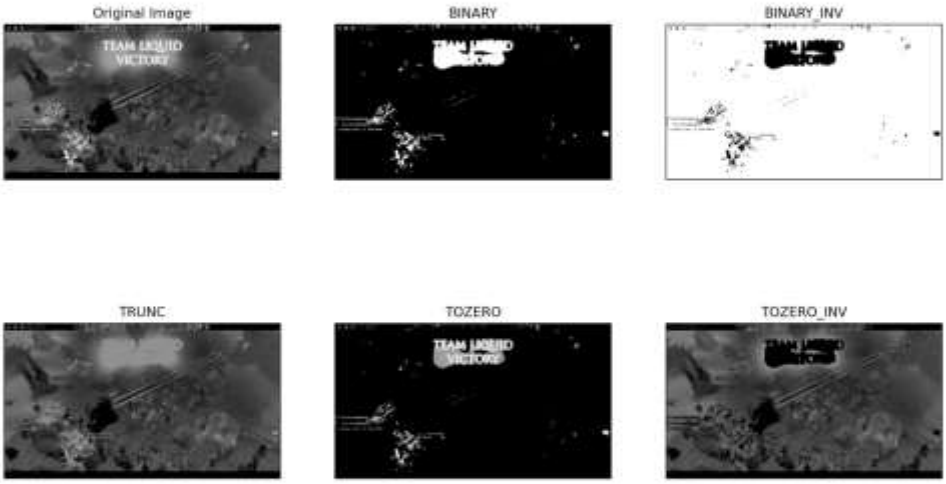
USER INTERFACES OF APPLICATION

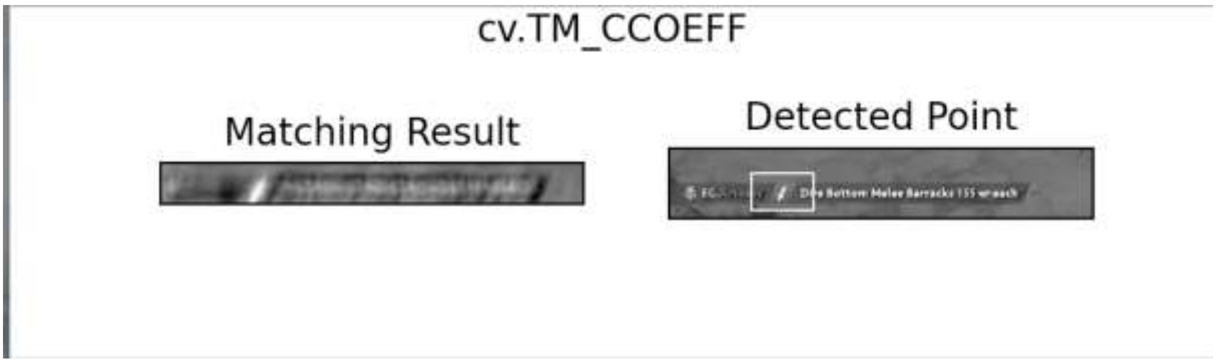


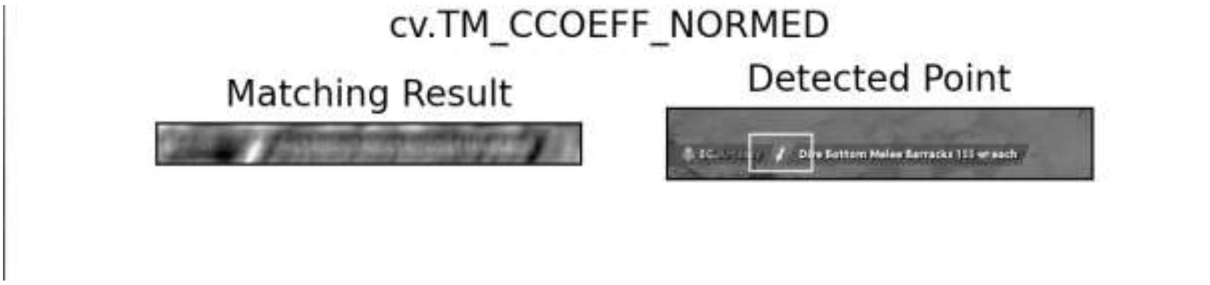


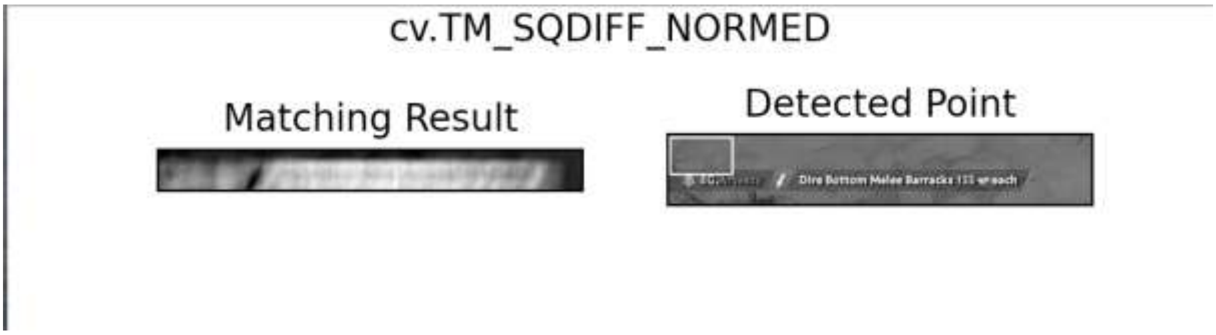
INTERMIDIATE RESULTS

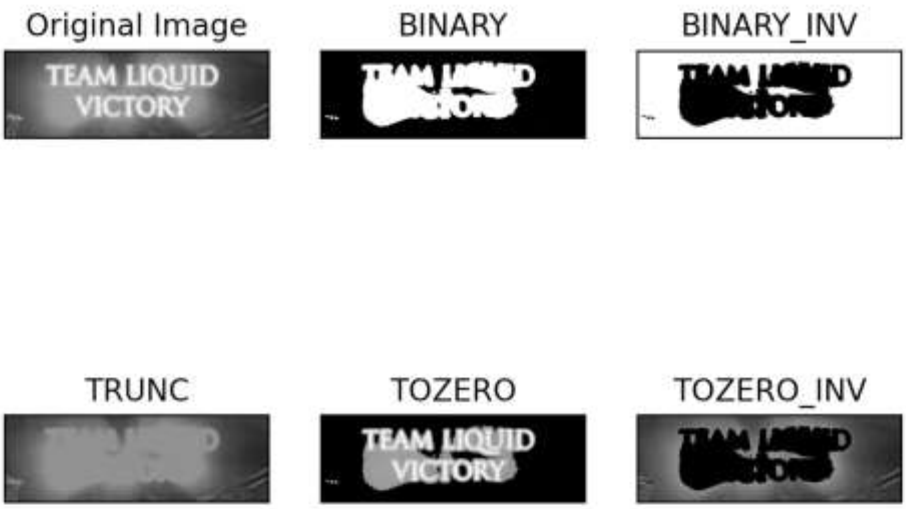



Original Image
BINARY
BINARY_INV
TRUNC
TOZERO
TOZERO_INV


cv.TM_CCOEFF
Matching Result
Detected Point


cv.TM_CCOEFF_NORMED
Matching Result
Detected Point


cv.TM_SQDIFF_NORMED
Matching Result
Detected Point


Original Image
TEAM LIQUID
VICTORY
BINARY
BINARY_INV
TRUNC
TOZERO
TEAM LIQUID
VICTORY
TOZERO_INV